AMENDMENTS TO THE SPECIFICATION:

Please amend the specification as follows.

Replace the second paragraph on p. 2 with the following paragraph:

U.S. Patent Application Ser. No. 09/409,260, entitled "METHODS AND APPARATUS FOR EXTRACTING ATTRIBUTES OF GENETIC MATERIAL," filed on the same date herewith by Jeffrey Saffer et al.;

Replace the first full paragraph on p. 14 with the following two paragraphs:

If the validation process determines that the data is sequence data, such as genome sequence data (step 312), the process determines whether the sequence data is in FastA file format (step 322) or whether the sequence data is in a SwissProt file format (step 324). An example FastA input file is provided in Appendix B. The operations and data associated with processing sequence data is discussed in more detail in U.S. Patent application serial no. 09/409,260, now issued as U.S. Patent No. 6,898,530, entitled "Methods and Apparatus for Extracting Attributes of Genetic Material" filed on the same day herewith by Jeffrey Saffer, et al. As stated in that patent, the steps for processing sequence data may comprise:

(i) comparing a sequence of each biopolymer material to a sequence of each other biopolymer material to provide respective comparison results;

(ii) arranging the comparison results in a square matrix indexed by the plurality of biopolymer materials; and

(iii) creating a high-dimensional context vector for at least one of the biopolymer materials based on a row or column of the square matrix.

If the sequence data is not in either of these formats, an error message is generated (step 320). If, however, the data is either a FASTA file (step 322) or a SwissProt file (step 324), the appropriate formats and delimiters, as discussed herein, are determined to be used for the respective FASTA file or SwissProt file (step 330). After the appropriate format/delimiters for the data type are determined (step 330), the corresponding format file/record delimiters are established (step 340). The format file/record delimiters specify the valid formats for reading the files and identifies the meta data files that are to be used for subsequent processing of the data set as discussed herein.

Replace with the bridging paragraph starting on page 23, at line 20, and ending on page 24, at line 10, with the following two paragraphs:

The visualizations discussed herein are based on high-dimensional context vector representations of the data. Thus, each type of data is represented in that manner. For purely numeric data, the vector representation is simply the values associated with each record attribute. For categorical data, the vector representation can be based on any method that translates categorical values or the distances between values as a number. For text data, the vector representation can be derived by latent semantic indexing as known to those skilled in the art or by related methods, such as described in U.S. patent application serial number No. 08/713,313, entitled "System for Information Discovery," filed on September 13, 1996 (now issued as U.S. Patent No. 6,772,170). As stated in that patent, the steps for processing text data may comprise:

a) semantically filtering a set of documents in a database to extract a set of semantic concepts, to improve an efficiency of a predictive relationship to its content, based on at least one of word frequency, overlap and topicality;

b) defining a topic set, said topic set being characterized as the set of semantic concepts which best discriminate the content of the documents containing them, said topic set being defined based on at least one of word frequency, overlap and topicality;

c) forming a matrix with the semantic concepts contained within the topic set defining one dimension of said matrix and the semantic concepts contained within the filtered set of documents comprising another dimension of said matrix;

d) calculating matrix entries as the conditional probability that a document in the database will contain each semantic concept in the topic set given that it contains each semantic concept in the filtered set of documents; and

e) providing said matrix entries as vectors to interpret the document contents of said database.

For sequence data, the context vector can be derived from any combination of numerical or categorical attributes of the sequence or by methods described herein. In addition, a user skilled in the art will recognize that the vectors created for each record do not have to be created from a single data type. Rather, the vectors can be created from mixed mode data, such as combined numeric and text data.